Volume 13 Number 21 1985 Nucleic Acids Research

The nucleotide sequence of the nitrogen regulation gene *ntrB* and the *glnA-ntrBC* intergenic region of *Klebsiella pneumoniae*

Stuart A.MacFarlane and Mike Merrick

AFRC Unit of Nitrogen Fixation, University of Sussex, Brighton BN1 9RQ, UK

Received 6 September 1985; Accepted 8 October 1985

ABSTRACT

The nucleotide sequence of the Klebsiella pneumoniae ntrB gene and the glnA-ntrBC intergenic region has been determined. NtrB encodes a 38,409 Dalton polypeptide with a potential DNA-binding domain between residues 67 and 86. This N-terminal domain may play a role in the co-operative control of ntr-regulated promoters by the ntrB and ntrC products. Mapping of in vivo transcripts with S1 nuclease identified three transcripts in the glnA-ntrBC intergenic region. Two transcripts originate upstream of glnA; one reading through into ntrBC and one terminating at a sequence resembling a rho-independent terminator between glnA and ntrBC. A third transcript originates from the ntrBC promoter which has a consensus binding site for the ntrC product in the -10 region. Comparison of the glnA-ntrBC intergenic sequences from K. pneumoniae, Escherichia coli and Salmonella typhimurium has identified a number of conserved features and some significant differences.

INTRODUCTION

Klebsiella pneumoniae and other related enteric bacteria (e.g. Escherichia coli and Salmonella typhimurium) are able to regulate various nitrogen assimilation enzymes in response to the level of fixed nitrogen in their environment (1). In recent years this regulation has been shown to be mediated by the products of three ntr (nitrogen regulation) genes which in K. pneumoniae play a central role in the expression of the nitrogen fixation (nif) genes (2,3,4).

Two of the ntr genes, ntrB and ntrC, are adjacent to glnA (the structural gene for glutamine synthetase, G.S.) and the third gene, ntrA, is unlinked. The ntrB and ntrC genes were initially described in S. typhimurium (5), ntrB coding for a 36 kDal polypeptide (NtrB) and ntrC coding for a 54 kDal polypeptide (NtrC). These genes were subsequently identified in K.

pneumoniae, E. coli and K. aerogenes (in the latter two organisms ntrA, ntrB and ntrC are referred to as glnF, glnL and glnG respectively) (2,6,7).

Transcription of ntrBC can be initiated either from one of two promoters preceding glnA, reading through this gene into ntrB and ntrC, or from a third promoter immediately preceding ntrB (8,9). During nitrogen limitation ntrBC transcription proceeds primarily from pglnA, whilst during nitrogen sufficiency transcription proceeds primarily from pntrB (8). Although transcription of glnA and ntrBC is coordinated by transcription initiating upstream of glnA, NtrB and NtrC are not synthesised in equimolar amounts to G.S. (2,8). The lower levels of NtrB and NtrC have been attributed to transcription termination within the glnA-ntrB intergenic region (10,11).

NtrC has been shown to be a bifunctional regulatory protein capable of repressing transcription from some promoters and of activating transcription from others (8,12). Examination of the DNA sequence of several promoters at which NtrC acts as a repressor has identified a homologous region with the consensus TGCACTANNNTGGTGCAA (9,11,13). This sequence is homologous to a number of known repressor binding sites and has been shown to bind purified NtrC in vitro (11,13,14).

The precise function of the ntrB product is unclear. Mutations in the gene produce a variety of phenotypes including low constitutive synthesis of G.S. (GlnR phenotype), high constitutive synthesis of G.S. (GlnC) and the suppression of glnD, glnB and ntrA mutations to allow growth without exogenous glutamine (1,15,16). These phenotypes have led to the hypothesis that NtrB is able to respond to signals (perhaps from the glnD and glnB products) concerning the nitrogen status of the cell and then to modulate the repressor or activator functions of NtrC accordingly (17). In this context it has been suggested that NtrB may form a complex with NtrC under certain conditions (5,15,18).

In order to study the role of NtrB in more detail and as a first step towards identifying specific domains in the protein we have determined the complete nucleotide sequence of the K. pneumoniae ntrB gene. We have also sequenced the glnA-ntrB

intergenic region and identified transcription initiation and termination sites within this sequence. Comparison with the analogous intergenic sequences from E. coli and S. typhimurium has revealed a number of conserved features and some notable differences.

MATERIALS AND METHODS

Cloning and DNA Sequencing

DNA carrying the ntrB sequence was derived from plasmids pMM12 (2) and pAM61 (8). To obtain sequence data between the KpnI site at 172 bp and the SalI site at 864 bp, this region was subjected to Bal31 digestion (leftward from the SalI site) to produce the two plasmids pSM2 and pSM3.

Restriction enzymes and DNA-modifying enzymes were obtained from commercial sources and used according to the manufacturers' instructions. Dideoxy sequencing reactions were carried out using clones prepared in M13 mp8, mp11, mp19 and pEMBL8$^+$ vectors (19,20) with [α^{35}S]-dATP as the labelled nucleotide (21).

Computer analysis of sequence data employed programs described elsewhere (22,23).

RNA Extraction and S1 Nuclease Mapping

RNA for S1 nuclease transcription mapping was isolated from E. coli strain ET8894 (Δ rha glnA ntrB ntrC 1703::Mu cts, rbs, gyrA, lacZ::IS1, hut C_k^c)(24) carrying the plasmids pGE100 (glnA$^+$ ntrB$^+$C$^+$)(2) or pSM2 (glnA', ntrB-lacZ). Cells were grown overnight at 32°C in NFDM medium (25) containing glutamine at 200 µg/ml plus the appropriate antibiotic selection. The cells were centrifuged and resuspended in fresh NFDM containing glutamine at 200 µg/ml (-N), glutamine at 2 mg/ml plus ammonium sulphate at 2 mg/ml (+N), or Luria broth (26) plus glutamine at 200 µg/ml (LB) and grown for a further 4h at 32°C. RNA was extracted by the method of Aiba (27).

The preparation of strand-specific probes for S1 nuclease mapping and the conditions for hybridising probe DNA to mRNA have been described before (9). For these experiments two uniformly ^{32}P-labelled DNA probes derived from pSM2 were synthesised using M13 mp8 templates carrying either the 350 bp

EcoRI-XhoII fragment containing the glnA-ntrB intergenic region or the 465 bp EcoRI-BamHI fragment carrying the intergenic region and 145 bp of ntrB. The latter probe was shortened by restriction with PvuII to unambiguously locate the positions of transcripts protected by the probe against S1 nuclease digestion.

RESULTS

The Nucleotide Sequence of ntrB

Previous work in this laboratory showed that the K. pneumoniae ntrB gene is carried on the plasmids pAM61 and pMM12, the former plasmid encoding a truncated 34 kDal ntrB product and the latter plasmid encoding the C-terminal portion of ntrB together with the whole of ntrC (2,8). Our sequencing strategy was based on the restriction map of ntrB derived from these plasmids (8). A more detailed map derived from the sequence data, and details of the clones used for the sequencing are given in Fig. 1.

A sequence of 1408 bp was obtained, extending from the EcoRI site at the C-terminus of glnA to a HindII site at the N-terminal end of ntrC (Fig. 2). The data revealed two extensive open-reading frames. The first extends 129 bp from the leftmost EcoRI site and has been identified as the C-terminus of glnA by

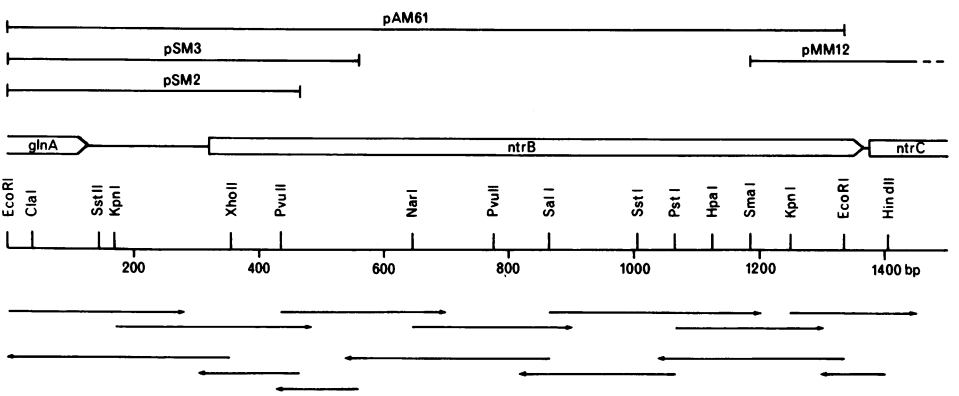
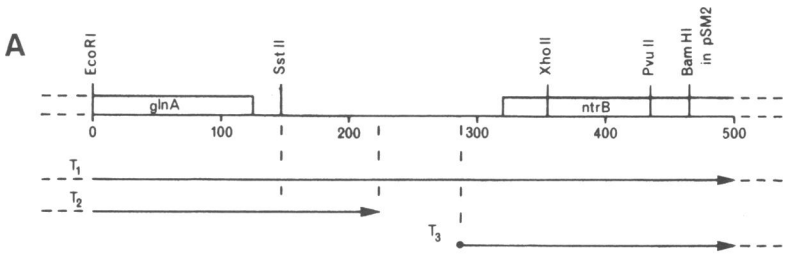
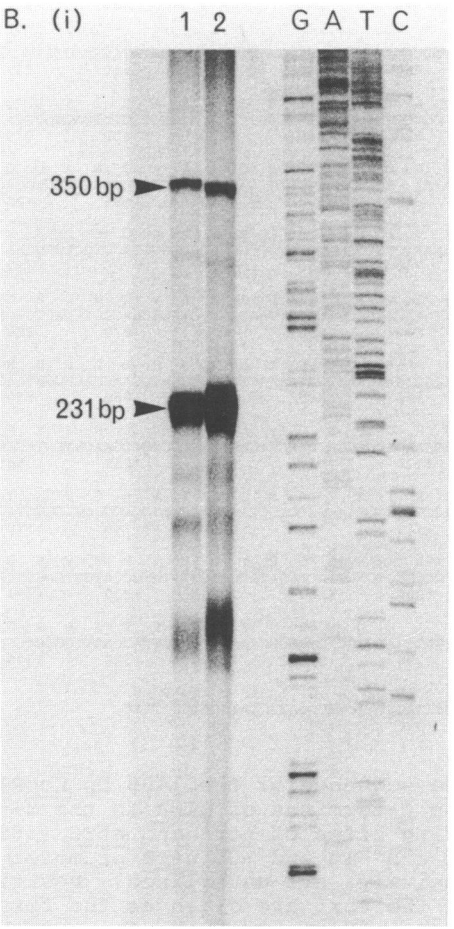


Fig. 1. Restriction map of the K. pneumoniae ntrB gene and flanking sequences. The extent of clones used for sequencing are shown above the restriction map and the extent of the nucleotide sequence obtained from each M13 clone is indicated by arrows below the map.

```
 E  F  L  T  A  G  G  V  F  T  N  D  A  I  D  A  Y  I  A  L  R  I  E  E  N  D  R  V  R  M
GAATTCCTGACCGCTGGCGGCGTATTCACCAACGACGCTATCGATGCTTACATCGCCCTGCGTATTGAAGAAAACGACCGCGTACGCATG
      10        20        30        40        50        60        70        80        90

 T  P  H  P  V  E  F  E  L  Y  Y  S  V  *
ACTCCGCACCCGGTTGAGTTTGAGCTGTACTACAGCGTCTAATTTTGCTGTAAATCCGCGGTATTCCGTGTTGTGGCGTGGGGTACCAAG
      100       110       120       130       140       150       160       170       180

GATGTTTTGTTGCCGTGGAAACTTTCAGCCCATCTCAGGATGGGCTTTTTTCTCCACCAACGCGCTGATCCCATGAGCTTTTTATTCATA
      190       200       210       220       230       240       250       260       270

                *                                                   M  A  T  G  T  L  P  D  A  G  Q  I  L  N   14
AAAAGCTATACTGCACTAAAATGGTGCACTCTTTTTCAGGAGACTGTCGAATGGCAACAGGCACACTGCCCGATGCTGGGCAGATCCTTA
      280       290       300       310       320       330       340       350       360

   S  L  I  N  S  I  L  L  V  D  D  D  L  A  V  H  Y  A  N  P  A  A  Q  Q  L  L  A  Q  S  S   44
ACTCTTTAATCAACAGTATCTTACTGGTCGATGACGACCTGGCGGTACATTACGCGAACCCGGCGGCGCAGCAGCTGCTGGCGCAAAGCT
      370       380       390       400       410       420       430       440       450

   R  K  L  F  G  T  P  L  P  E  L  L  S  Y  F  S  L  N  I  G  L  M  Q  E  S  L  A  A  G  Q   74
CGCGTAAATTATTCGGCACGCCGCTGCCGGAGCTGTTGAGCTATTTCTCGCTGAATATCGGCCTCATGCAAGAGAGCCTCGCTGCGGGTC
      460       470       480       490       500       510       520       530       540

   G  F  T  D  N  E  V  T  L  V  I  D  G  R  S  H  I  L  S  L  T  A  Q  R  L  P  E  G  Y  I  104
AGGGTTTTACCGATAACGAAGTGACGCTGGTTATTGATGGCCGCTCCCACATCCTTTCCCTGACGGCGCAACGCCTGCCTGAAGGCTATA
      550       560       570       580       590       600       610       620       630

   L  L  E  M  A  P  M  D  N  Q  R  R  L  S  Q  E  Q  L  Q  H  A  Q  Q  I  A  A  R  D  L  V  134
TCCTGCTGGAAATGGCGCCGATGGATAATCAGCGTCGCCTGAGCCAGGAGCAGCTACAGCACGCGCAGCAGATTGCCGCACGTGACTTAG
      640       650       660       670       680       690       700       710       720

   R  G  L  A  H  E  I  K  N  P  L  G  G  L  R  G  A  A  Q  L  L  S  K  A  L  P  D  P  A  L  164
TGCGTGGTCTGGCCCACGAGATTAAAAATCCGCTGGGCGGCCTGCGCGGCGCGGCGCAGCTGCTGAGCAAAGCGCTGCCCGACCCTGCGC
      730       740       750       760       770       780       790       800       810

   M  E  Y  T  K  V  I  I  E  Q  A  D  R  L  R  N  L  V  D  R  L  L  G  P  Q  H  P  G  M  H  194
TAATGGAATATACCAAAGTCATTATTGAACAGGCCGACCGGTTAAGAAATCTGGTCGACCGTCTACTGGGACCACAGCATCCCGGAATGC
      820       830       840       850       860       870       880       890       900

   V  T  E  S  I  H  K  V  A  E  R  V  V  K  L  V  S  M  E  L  P  D  N  V  K  L  V  R  D  Y  224
ACGTAACAGAAAGCATCCACAAGGTCGCTGAGCGAGTGGTGAAGCTGGTGTCGATGGAACTACCGGATAACGTCAAGCTGGTGCGCGACT
      910       920       930       940       950       960       970       980       990

   D  P  S  L  P  E  L  P  H  D  P  D  Q  I  E  Q  V  L  L  N  I  V  R  N  A  L  Q  A  L  G  254
ATGACCCCAGCCTGCCAGAGCTCCCGCACGACCCGGATCAAATTGAACAAGTGCTGCTGAACATCGTACGTAACGCGCTGCAGGCGCTGG
      1000      1010      1020      1030      1040      1050      1060      1070      1080

   P  E  G  G  E  I  T  L  R  T  R  T  A  F  Q  L  T  L  H  G  V  R  Y  R  L  A  A  R  I  D  284
GACCGGAGGGCGGAGAGATCACTCTGCGCACCCGGACGGCCTTCCAGTTAACCCTGCACGGCGTGCGCTACCGCCTCGCTGCGCGTATAG
      1090      1100      1110      1120      1130      1140      1150      1160      1170

   V  E  D  N  G  P  G  I  P  S  H  L  Q  D  T  L  F  Y  P  M  V  S  G  R  E  G  G  T  G  L  314
ACGTTGAAGATAACGGCCCGGGAATCCCGTCTCACTTACAGGATACGCTGTTTTATCCCATGGTGAGTGGACGCGAAGGCGGTACCGGAC
      1180      1190      1200      1210      1220      1230      1240      1250      1260

   G  L  S  I  A  R  S  L  I  D  Q  H  S  G  K  I  E  F  T  S  W  P  G  H  T  E  F  S  V  Y  344
TCGGGCTGTCTATCGCCCGCAGCCTTATCGACCAGCACTCCGGCAAAATTGAATTTACCAGTTGGCCAGGGCATACCGAATTCTCGGTGT
      1270      1280      1290      1300      1310      1320      1330      1340      1350

   L  P  I  R  K  *           M  Q  R  G  I  A  W  I  V  D
ACCTGCCTATTCGGAAGTAGAGGTGTTTATGCAACGAGGGATAGCATGGATCGTTGAC
      1360      1370      1380      1390      1400
```

Fig. 2. Nucleotide sequence of the 1408 bp EcoRI-HindII fragment extending from the C-terminus of glnA to the N-terminus of ntrC. The ribosome-binding sites of ntrB and ntrC (308 bp and 1369 bp respectively) and the -10 and -35 ntrB promoter sequences (277 bp and 253 bp respectively) are underlined. Positions of restriction sites detailed in the text are given as the first bp of the recognition sequence. Numbers at the far right of the figure indicate amino acid residues of NtrB. The transcription start site for ntrB is marked by the asterisk at 288 bp.


A
EcoR I
Sst II
Xho II
Pvu II
Bam HI in pSM2
glnA
ntrB
0
100
200
300
400
500
T1
T2
T3


B. (i)
1 2
G A T C
350 bp
231 bp

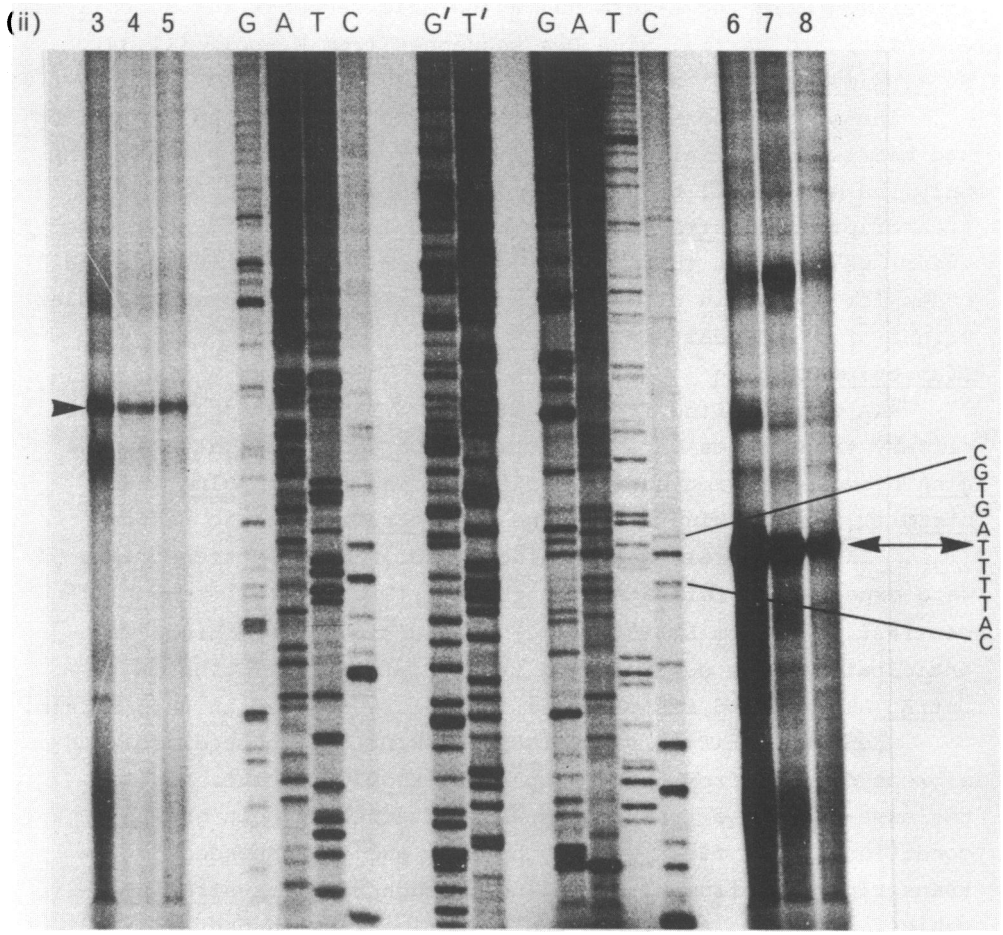


Fig. 3. S1 nuclease mapping of transcripts in the glnA-ntrB intergenic region. A) Schematic representation showing relevant region, restriction sites and proposed transcriptional organisation. B) Autoradiographs of protected fragments obtained by S1 nuclease digestion of RNA-DNA hybrids (i) RNA extracted from ET8894 (pGE100) grown in either high nitrogen medium (+N), lane 1; or low nitrogen medium (-N), lane 2; and probed with a 350 bp EcoRI-XhoII fragment. Lanes marked G, A, T, C are dideoxy sequencing products used to size the transcripts. (ii) RNA extracted from ET8894 (pSM2) grown in LB (lanes 3 and 6), +N (lanes 4 and 7) or -N (lanes 5 and 8) and probed with either the 350 bp EcoRI-XhoII fragment (lanes 3, 4, 5) or the 430 bp EcoRI-PvuII fragment (lanes 6, 7, 8) from pSM2. Lanes marked G, A, T, C are dideoxy sequencing products cut with BamHI or PvuII respectively to generate 5' termini equivalent to those of the probe. Lanes marked G' and T' are dideoxy sequencing products uncut by any restriction enzyme.

comparison with the C-terminal amino acid sequence for E. coli G.S. (28) and partial glnA DNA sequences from E. coli (11,34), S. typhimurium (29) and Anabaena 7120 (30).

The second open-reading frame extends from 321 bp to 1367 bp and has two potential ATG initiation codons at 321 bp and 291 bp. Only the ATG at 321 bp has a good ribosome binding site and as transcription of ntrB has been shown to initiate at 288 bp (see later) we conclude that ntrB extends from 321 bp to 1367 bp and codes for a protein of 349 amino acids with a predicted molecular weight of 38,409 daltons.

Transcript Mapping

RNA was purified from two strains. In ET8894 (pGE100) the primary transcripts were predicted to be those initiated in the glnA promoter region and either proceeding through glnA into ntrBC (T_1) or terminating in the glnA-ntrB intergenic region (T_2). In this strain only a minority of the ntrBC transcripts were expected to initiate a pntrBC (T_3)(Fig. 3A)(8). By contrast, RNA from ET8894 (pSM2) hybridising to the probe was anticipated to be derived from transcription initiating at pntrBC as glnA and its promoter region are deleted in this strain.

When RNA purified from ET8894 (pGE100) was hybridised with a probe derived from the 350 bp EcoRI-XhoII fragment, two S1 nuclease-protected transcripts were detected in both growth conditions. The first was 350 bp long and corresponded to transcript (T_1) from glnA reading through the glnA-ntrB intergenic region into ntrB. Control experiments with unhybridised probe DNA digested by S1 nuclease showed that the 350 bp molecule was not simply the result of incomplete digestion by the enzyme. A second transcript of 231 bp was also detected by this probe (see Fig. 3B(i)).

When the same RNA was hybridised with the EcoRI-PvuII probe, derived from pSM2, the 231 bp transcript was retained. When the RNA was hybridised with the EcoRI-XhoII probe, which had been cleaved with SstII before hybridisation, the 231 bp transcript was replaced by a 147 bp transcript. These results show that the 231 bp transcript (T_2) spans a region from the leftmost EcoRI site to a point within the glnA-ntrB intergenic region. Analysis of the DNA sequence at 231 bp downstream from the EcoRI

Table 1. Comparison of the NH_2-terminal amino acid sequence of ntrB from K. pneumoniae, E. coli and S. typhimurium (non-homologous amino acids are underlined).

K. pneumoniae	M A T G T L P D A G Q I L N S L I N S I L L	
E. coli	M A T G T Q P D A G Q I L N S L I N S I L L	(ref. 11)
S. typhimurium	M A S G I Q P D A G Q I L N S L I N S V L L	(ref. 29)

site shows that the endpoint of the transcript coincides with a run of six thymidine residues adjacent to a potential stem/loop-forming sequence (see Discussion).

To identify the site of transcription initiation for ntrB, RNA from ET8894 (pSM2) was probed with ^{32}P-labelled DNA from the same plasmid. The probe was cleaved with either BamHI or PvuII before hybridisation and the reaction products were sized by running them next to dideoxy sequencing ladders of the probe also cut with BamHI or PvuII. In each experiment a single transcript (T_3) was detected which was shown to initiate at the first of a run of four adenine residues situated 33 bp upstream of the start of ntrB (see Fig. 3B(ii)).

DISCUSSION

DNA Sequence of ntrB

The DNA sequence obtained for ntrB and the glnA-ntrB intergenic region confirms previous studies made in this laboratory. The calculated molecular weight of 38,409 daltons is in agreement with the previous estimate of 36 kDal derived from 2D PAGE analysis of polypeptides synthesised in an in vitro transcription/translation system (2). Plasmid pAM61 has been shown to code for only 339 of the 349 amino acids in NtrB thus explaining the smaller polypeptide synthesised from this plasmid in an in vitro transcription/translation system (8).

The N-terminal amino acid sequence of K. pneumoniae ntrB shows greater than 82% homology with the partial sequences available for E. coli and S. typhimurium (Table 1). This finding suggests that the ntrB gene in these organisms is probably highly conserved.

Table 2. Alignment of NtrB with other site-specific DNA-binding proteins showing the 20 amino acid region of homology.

	67		70						75						80					85		
ntrB	Q	E	S	L	A	A	G	Q	G	F	T	D	N	E	V	T	L	V	I	D		
λrep (33)	Q	E	S	V	A	D	K	M	G	M	G	Q	S	G	V	G	A	L	F	N		
λcro (16)	Q	T	K	T	A	K	D	L	G	V	Y	Q	S	A	I	N	L	A	I	H		
434 rep (17)	Q	A	E	L	A	Q	K	V	G	T	T	Q	Q	S	I	E	Q	L	E	N		
434 cro (19)	Q	T	E	L	A	T	K	A	G	V	K	Q	Q	S	I	Q	L	I	E	N		
P22 rep (21)	Q	A	A	L	G	K	M	V	G	V	S	N	V	A	I	S	Q	W	Q	R		
P22 cro (13)	Q	R	A	V	A	K	A	L	G	I	S	D	A	A	V	S	Q	W	K	E		

Residues that are identical in NtrB and in one or more of the other sequences are underlined. Numbers above the sequence refer to amino acid positions in NtrB and numbers in brackets refer to the position of the first amino acid shown in each protein. Alignments of the λ, 434 and P22 sequences are from Pabo and Sauer (31).

The close linkage between ntrB and ntrC has previously made the accurate mapping of mutations in these genes very difficult (15,16) and further complexity arises from the polar nature of certain ntrB mutations onto ntrC (15). As far as we are aware only one ntrB mutation has as yet been defined precisely. This mutation was created by deletion of a 480 bp ClaI-SalI fragment from the N-terminus of the E. coli ntrB gene, blunting the ends with DNA polymerase and religating the remaining parts of the gene (16). Comparison of the E. coli restriction map of ntrB with the DNA sequence of K. pneumoniae ntrB suggests that a potential ClaI site ATCGAT is replaced by the sequence GTCGAT at 387 bp in the K. pneumoniae sequence. Deletion of 476 bp between this site and the SalI site at position 864 would allow the downstream sequence to remain in frame and thus the phenotype of this mutant, which responds slowly to changes in the nitrogen status of the cell may be due to a truncated but partially active ntrB product.

Sequence Homology between ntrB and other DNA-binding Proteins

Crystallographic studies of lambda repressor, lambda Cro and E. coli CAP protein indicate that these DNA-binding proteins recognise DNA in a similar manner. Many of the DNA contacts are

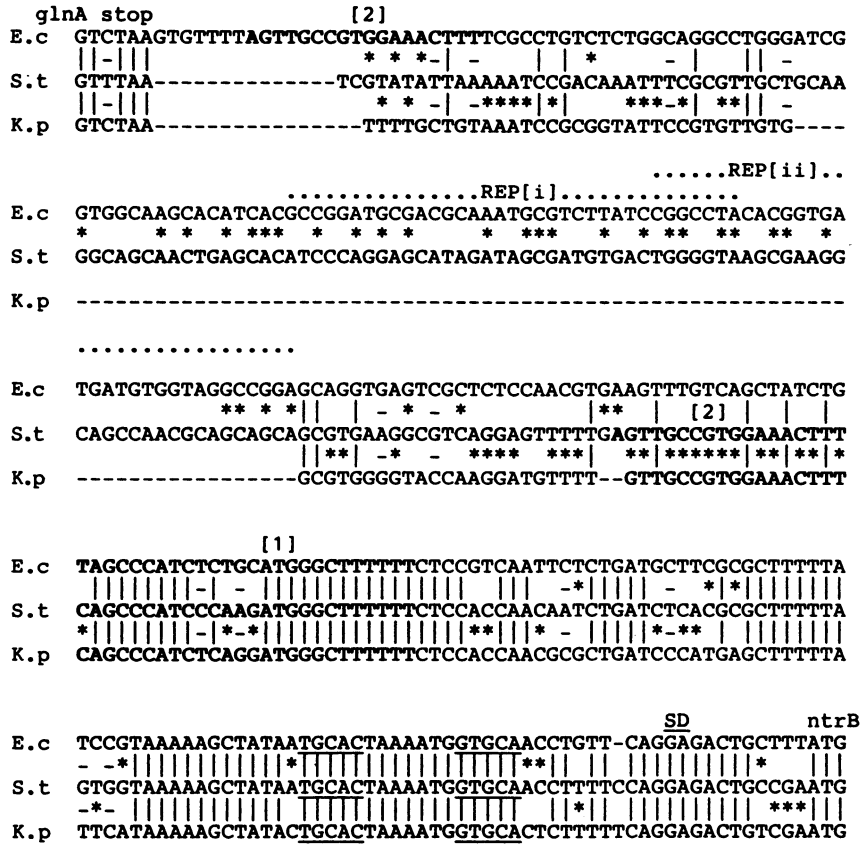


Fig. 4. Alignment of the glnA-ntrB intergenic sequence of E. coli (E.c), S. typhimurium (S.t) and K. pneumoniae (K.p). Space bars have been inserted within the nucleotide strings to maximise homology between the three sequences. The non-coding strand only is shown. The figures [1] and [2] overlie Loop 1 and Loop 2 respectively (see text) which are presented in bold type. REP [i] indicates the position of the Repetitive Extragenic Palindromic sequence (see text) present in the non-coding strand of the E. coli sequence. REP [ii] is present in the coding strand of the E. coli sequence. Homologies between the three sequences are highlighted as follows; | homology between in all three sequences, * homology between adjacent sequences, - homology between non-adjacent sequences. The binding site for NtrC is underlined. The Shine-Dalgarno (SD) site is indicated. The termination codon for glnA and the initiation codon for ntrB are indicated.

made by two α-helices linked by a tight turn and amino acid sequence homologies between these regions of lambda Rep, Cro and CAP and other DNA-binding proteins suggest that many site-specific DNA binding proteins have similar bihelical binding domains (31).

A computer search of the NtrB amino acid sequence using the program PATSCAN (32) identified a region between residues 67 and 86 with extensive homology to the known or predicted DNA-binding domains of several other site-specific DNA-binding proteins (J. Wootton, pers. comm.). The NtrB sequence resembles other DNA-binding sequences in several critical respects (Table 2). (i) The glycine residue, at position 75, which is highly conserved at the analogous position in other DNA-binding proteins as it plays an important role in the formation of the turn between the two α-helices. (ii) The presence of the highly conserved glutamine and alanine residues at positions 67 and 71; and (iii) the presence of hydrophobic residues at positions 70, 81 and 84. Based on these comparisons it seems very likely that an N-terminal domain of NtrB is involved in binding of the protein to DNA. Previous genetic analysis of mutations in the *ntrB* gene of *E. coli* identified the N-terminal region of the gene as being involved in a negative regulatory function. This region was identified by mapping *ntrA*-suppressor mutations and mutations resulting in elevated derepression of G.S. (16). Furthermore previous studies on regulation of the *K. pneumoniae* *glnA ntrBC* operon (8,9) suggested that an *ntrB-lacZ* translational fusion containing the N-terminal 800 bp of *ntrB* retained some negative regulatory function. These results are consistent with an N-terminal domain of NtrB, which would contain the predicted DNA-binding site, having a role in the repression of *glnA* transcription.

The *ntrC* product has already been shown to be a DNA-binding protein (13,15,33) and it has been proposed that NtrB can complex with NtrC to modulate or modify the regulatory activity of NtrC (15,18). NtrB is required in addition to NtrC both to activate and repress transcription from the tandem *glnA* promoters in *K. pneumoniae* (8,9) and the presence of a potential DNA-binding domain in NtrB may indicate that this cooperativity is achieved

by the concerted DNA-binding of an NtrB/NtrC complex. Further studies are required to determine whether NtrB can bind sequences in the glnA promoter region either in the absence or the presence of NtrC.

The glnA-ntrB Intergenic Region

The glnA-ntrB intergenic region in K. pneumoniae although significantly shorter (191 bp) than in E. coli (288 bp) and S. typhimurium (275 bp), shares several defined homologous sequences with these organisms (Fig. 4). Alignment of the intergenic regions such that their 3' ends coincide reveals extensive conservation of both the order and spacing of the homologous features. The K. pneumoniae glnA-ntrB intergenic region shares 80% sequence homology in the distal 114 bp when compared with E. coli and 77% homology in the distal 157 bp when compared with S. typhimurium (see Fig. 4).

The functional ntrB promoter sequence has been identified by S1 nuclease mapping of the RNA transcript initiating within the glnA-ntrB intergenic region. The promoter has a -10 sequence (TATACT) with a good match to the consensus (TATAAT) but there is only a poor match in the -35 region (ATGAGC) with the consensus TTGACA. The site of ntrB transcription initiation is identical in K. pneumoniae and E. coli and in both these organisms this site is bounded by a sequence previously identified as the binding site for the repressor form of NtrC (9,11,15). This sequence is also present in the analogous position in the S. typhimurium intergenic region (29). It should be noted that in the upstream promoter (RNA2) of glnA, the NtrC binding site is slightly upstream of the transcription initiation site and actually surrounds the -10 region although in both the glnA and ntrBC promoters binding of NtrC would effectively inhibit RNA polymerase binding. By contrast, the NtrC binding site described in the S. typhimurium dhuA promoter region (13) lies 116 bp upstream of the site of transcription initiation and, as there is presently no genetic evidence for repression of dhuA transcription by NtrC, the role of this binding site in dhuA regulation is unclear.

Apart from the 114 bp region around the ntrB promoter two other regions are highly conserved in K. pneumoniae, E. coli and

S. typhimurium. Each of these regions include sequences of inverted symmetry with the potential to form hairpin structures in the mRNA.

The first hairpin (position 189 in K. pneumoniae) which we have termed loop 1, is highly conserved in both sequence and position and has a standard free energy of formation of -18.4 kcal. Other authors have suggested that this stem/loop structure may be a rho-independent terminator and could affect transcription through the glnA-ntrB intergenic region (10,11). Our transcript mapping data confirms this hypothesis and show unambiguously that in K. pneumoniae a significant proportion of transcripts passing through the glnA-ntrB intergenic region terminate at loop 1. Experiments in this laboratory using glnA- and ntrB-β-galactosidase fusions have indicated that as much as 80% of transcripts from pglnA terminate upstream of the ntrB promoter (8), and a somewhat similar figure has been reported for E. coli (6). Our transcript mapping data is in agreement with these figures (see Fig. 3b(i)).

The second hairpin, loop 2, is positioned immediately upstream of loop 1 in both K. pneumoniae and S. typhimurium (see Fig. 4) and whilst it is not itself a potentially stable mRNA structure it might cooperate with loop 1 in effecting transcription termination. However, in E. coli, although conserved in sequence, loop 2 is situated at the upstream end of the intergenic region and separated from loop 1 by a region containing two copies of an element termed REP (Repetitive Extragenic Palindromic sequence) (34,35,36). Although no conserved REP sequences are present in S. typhimurium 21 bp out of 60 bp in this region are homologous to the E. coli sequences (Fig. 4) suggesting that S. typhimurium may originally have carried REP sequences at this position. It was previously proposed that as no major differences have been detected in the regulation of expression of the glnA ntrBC operons of E. coli and S. typhimurium, the REP sequence, which is absent from S. typhimurium, plays no part in such regulation (11). However the conservation of loops 1 and 2 in all 3 organisms suggests that both sequences may have a role in controlling transcription through the glnA-ntrB intergenic region and that the juxtaposition of loops 1 and 2 in

K. pneumoniae and S. typhimurium may compensate for the absence of some control mediated by the REP sequences present in E. coli.

The availability of the K. pneumoniae ntrB sequence will now allow the construction of defined mutations in ntrB and a detailed analysis of the functions of NtrB and its interaction with other components of the regulatory system of bacterial nitrogen assimilation.

ACKNOWLEDGEMENTS

We thank Ray Dixon, Martin Buck and Martin Drummond for valuable discussions and John Wootton for expert help in computer analysis of the DNA sequence. S. MacFarlane acknowledges a research assistantship from the A.F.R.C. Finally we thank Beryl Scutt for typing the manuscript.

REFERENCES

1. Magasanik, B. (1982) Ann.Rev.Genet., 16, 135-168.
2. Espin, G., Alvarez-Morales, A., Cannon, F., Dixon, R. and Merrick, M. (1982) Mol.Gen.Genet., 186, 518-524.
3. Merrick, M.J. (1982) Nature 297, 362-363.
4. Dixon, R.A. (1984) J.Gen.Microbiol., 130, 2745-2755.
5. McFarland, N., McCarter, L., Artz, S. and Kustu, S. (1981) Proc.Natl.Acad.Sci.(U.S.A.) 78, 2135-2139.
6. Pahel, G., Rothstein, D.M. and Magasanik, B. (1982) J.Bacteriol., 150, 202-213.
7. Rothman, N., Rothstein, D.M., Foor, F. and Magasanik, B. (1982) J.Bacteriol., 150, 221-230.
8. Alvarez-Morales, A., Dixon, R. and Merrick, M. (1984) EMBO J. 3, 501-507.
9. Dixon, R. (1984) Nucl.Acids Res., 12, 7811-7830.
10. Krajewska-Grynkiewicz, K. and Kustu, S. (1984) Mol.Gen. Genet. 193, 135-142.
11. Ueno-Nishio, S., Mango, S., Reitzer, L.J. and Magasanik, B. (1984) J.Bacteriol., 160, 379-384.
12. Drummond, M., Clements, J., Merrick, M. and Dixon, R. (1983) Nature 301, 302-307.
13. Ames Ferro-Luzzi, G. and Nikaido, K. (1985) EMBO J., 4, 539-547.
14. Hawkes, T., Merrick, M. and Dixon, R. (1985) (submitted for publication).
15. MacNeil, T., MacNeil, D. and Tyler, B. (1982) J.Bacteriol., 150, 1302-1313.
16. Chen, Y.-M., Backman, K. and Magasanik, B. (1982) J. Bacteriol., 150, 214-220.
17. Backman, K.C., Chen, Y.-M., Ueno-Nishio, S. and Magasanik, B. (1983) J.Bacteriol., 154, 516-519.
18. Wei, G.R. and Kustu, S. (1981) Mol.Gen.Genet., 183, 392-399.
19. Messing, J. (1983) in Methods in Enzymology, Wu, R., Grossman, L. and Moldave, K. eds., Vol. 101, pp.20-77,

Academic Press, New York.
20. Dente, L., Cesareni, G. and Cortese, R. (1983) Nucl. Acids Res., 11, 1645-1655.
21. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc.Natl. Acad.Sci.(U.S.A.) 80, 3963-3965.
22. Staden, R. (1980) Nucl. Acids Res., 8, 3673-3694.
23. Staden, R. (1984) Nucl. Acids Res., 12, 521-538.
24. MacNeil, D. (1981) J.Bacteriol., 146, 260-268.
25. Dixon, R., Kennedy, C., Kondorosi, A., Krishnapillai, V. and Merrick, M. (1977) Mol.Gen.Genet., 157, 189-198.
26. Kennedy, C. (1977) Mol.Gen.Genet., 157, 199-204.
27. Aiba, C., Adhya, S. and de Crombrugghe, B. (1981) J.Biol. Chem., 256, 11905-11910.
28. Ginsburg, A. (1972) Adv.Protein Chem., 26, 1-79.
29. Hanau, R., Koduri, R.K., Ho, N. and Brenchley, J.E. (1983) J.Bacteriol., 155, 82-89.
30. Turner, N.E., Robinson, S.J. and Haselkorn, R. (1983) Nature 306, 337-342.
31. Pabo, C.O. and Sauer, R.T. (1984) Ann.Rev.Biochem., 53, 293-321.
32. Wootton, J.C., Howland, C.J. and Dix, N.I.M. (in preparation).
33. Reitzer, L.J. and Magasanik, B. (1983) Proc.Natl.Acad.Sci. (U.S.A.) 80, 5554-5558.
34. Rocha, M., Vázquez, M., Garciarrubio, A. and Covarrubias, A.A. (1985) (in press).
35. Higgins, C.F., Ames Ferro-Luzzi, G., Barnes, W.M., Clement, J.M. and Hofnung, M. (1982) Nature 298, 760-762.
36. Stern, M.J., Ames Ferro-Luzzi, G., Smith, N.H., Robinson, E.C. and Higgins, C.F. (1984) Cell 37, 1015-1026.